

On the Equivalence Between a Minimal Codomain Cardinality Riesz Basis Construction, a System of Hadamard–Sylvester Operators, and a Class of Sparse, Binary Optimization Problems

James D. B. Nelson, *Member, IEEE*

Abstract—Piecewise, low-order polynomial, Riesz basis families are constructed such that they share the same coefficient functionals of smoother, orthonormal bases in a localized indexing subset. It is shown that a minimal cardinality basis codomain can be realized by inducing sparsity, via ℓ_1 regularization, in the distributional derivatives of the basis functions and that the optimal construction can be found numerically by constrained binary optimization over a suitably large dictionary. Furthermore, it is shown that a subset of these solutions are equivalent to a specific, constrained analytical solution, derived via Sylvester-type Hadamard operators.

Index Terms—Riesz bases, basis construction, Fourier series, ℓ_p regularization, sparsity basis selection.

I. INTRODUCTION

THE Riesz basis framework affords a convenient, flexible, and constructive approach to the design of biorthogonal bases. We are here interested in the construction of piecewise, low-order polynomial, biorthonormal basis families which approximate the behavior of smoother, orthonormal bases.

A direct motivating example is the idea of ‘approximate signal processing’, championed by Nawab and Dorken [16]. Generally, this principle advocates an iterative refinement approach to signal processing by balancing system performance and available resources.

In signal processing, the use of binary approximations to the Fourier basis has been well studied [7], [10], [13]. This has spurred interest in the natural extension to approximate filtering applications [8], [20] as well as recent, specialized applications such as hardware-based pattern recognition [15].

Moving beyond piecewise constant to piecewise, low-order polynomial, and using an electronics context, Wei [27] argued the case for using square, triangular, and trapezoidal functions

as approximate alternatives to the Fourier basis because they are very easily generated by high-frequency pulse counters. Wei also established the corresponding dual bases by way of the Dirichlet inverse—the general principle of which is discussed by Hedenmalm *et al.* [9] in the Riesz basis setting.

In the full sense of the approximate signal processing paradigm, one may go further and design a quantized version of an orthonormal (e.g., Fourier) transform which can be iteratively refined to become ever more faithful to the original basis in a dynamic manner as system resources allow [18]. A structural health monitoring application of this ‘approximate frequency analysis’ is discussed by Penny *et al.* [21]. More generally, and citing the lack of trade-offs between visual quality and battery life in portable video players, Anastasia and Andreopoulos introduce an image processing software architecture based on the incremental refinement principle [1].

Perhaps less obvious examples of low-order basis design include: the decomposition of generalized cardinal series, or classical sampling theorems, into parallel realizations of multi-channel, multisampling rate theorems [17]; and, partly inspired by the lifting scheme [23], constructions which enhance the filter characteristic of existing wavelet bases by superposing translated and dilated versions of low-order wavelets [19].

Framing our construction in the Riesz basis setting, it will be shown that a minimal cardinality basis codomain can be realized by inducing sparsity in the distributional derivatives of the basis functions and that the optimal construction can be found numerically by constrained binary optimization. Furthermore, since the energy of the approximating bases are typically concentrated on local spectral intervals, ℓ_1 regularization is applied to the cost function, between the Riesz basis and orthonormal basis, to ensure that the approximating Riesz basis behaves like the orthonormal basis in a localized region of the spectral domain.

Although its origins are certainly not new, the use of sparse regularization to drive basis selection is a fairly timely topic, buoyed by the early work on basis pursuit ideas explored by Chen *et al.* [6]. In a broader sense, Fourier, wavelet basis, or dictionary construction also impacts the field of approximation theory; an excellent and detailed account of which is given by Candès [4]. More recently, Candès *et al.* offered a framework to accommodate signals which are sparse in redundant dictionaries [5]. Aside from renewing interest in existing redundant dictionaries, ranging from the very redundant mega-dictionaries of Chen *et al.* [6] to Kingsbury’s optimally redundant dual-tree

Manuscript received October 26, 2013; revised April 24, 2014 and July 20, 2014; accepted July 20, 2014. Date of publication August 01, 2014; date of current version September 04, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Antonio Napolitano.

The author is with the Department of Statistical Science, University College London, London CB21PZ, U.K. (website: <http://www.ucl.ac.uk/statistics/people/jamesnelson>).

This paper has supplementary downloadable multimedia material available at <http://ieeexplore.ieee.org> provided by the authors. This files therein reproduce a superset of Fig 5. This material is 7 KB in size.

Digital Object Identifier 10.1109/TSP.2014.2345346

complex wavelet frames [12], [22], this naturally prompts questions regarding what constitutes a ‘good’ redundant dictionary and, given data, how to infer the ‘best’ basis.

Although some of the main computational thrust in this work is inspired by ℓ_1 regularization for basis selection, our goal is comparatively more elementary. Instead of adapting bases to sparse representations of data and sampling paradigms, we focus on approximating fixed bases. That the ideas of regularization and sparsity remain relevant in this context is testament to the flexibility of the Riesz basis setting, whereby sampling operator construction is equivalent to basis construction by suitable changes or reinterpretations of the underlying Hilbert space. In short, similar optimization techniques that are used to develop optimal strategies for sparse sampling, reconstruction, and best basis selection (given data) can also be used to construct fixed bases with certain desirable properties, subject to a sparse, optimal, set of parameters.

For fixed bases, the problem of constructing filters with minimal cardinality has recently been invigorated by the upsurge in popularity of sparsity inducing norms. Baran *et al.* [3] proposed the use of ℓ_1 regularization, along with iterative ‘thinning’ to construct optimally sparse finite impulse response filters, constrained by frequency characteristic conditions. Moreover, filters with small dynamic range, in particular those endowed with a power-of-two-many coefficients, have also received interest. Li *et al.* [14] motivate their construction by observing that power-of-two VLSI filters can yield significant savings in the required chip area.

This paper is organised as follows. In Section II, we present some requisite concepts from Hilbert spaces, place our problem in a general Riesz basis setting, and fix notation. Section III marries, and then contrasts, two disparate constructions under this framework. The first, a reinterpretation of a Riesz basis construction that was originally used to form a generalized, sampling theorem, is motivational (Example 3.7); it is shown that the second construction, a uniformly quantized basis used in approximate signal processing, is intrinsically superior for our purposes in that it is guaranteed to have a smaller dynamic range (Definition 3.12 together with Example 3.18). Unlike previous work, the uniformly quantized basis is placed in the Riesz basis setting. We build on this result in Section IV where we prove and demonstrate the relationship between the optimal minimal cardinality, piecewise constant, uniform, construction and a constrained binary optimization problem with ℓ_1 regularization (Corollary 4.1 and Theorem 4.2). In particular, it is shown that vectors generated from Hadamard-Sylvester operators solve a class of Fourier-type binary optimization problems (Corollary 4.3). The framework is extended in Section V, in two different ways, to higher order polynomial constructions. Numerical results in Section VI: (i) corroborate the fact that the Hadamard-Sylvester solutions solve the reported class of binary optimization problems; (ii) show, at least for low-order cases, that the Hadamard-Sylvester solutions are optimal in that the numerical solutions do not perform any better; and (iii) show that the Hadamard-Sylvester solutions also solve a broader class of binary optimization problems.

II. PRELIMINARIES

In the interests of generality and convenience, it is instructive to place our development in the context of Hilbert spaces,

with particular attention on Riesz bases. The following declares many of the basic objects used throughout, introduces notation, and is also intended as motivation. The section concludes with a brief preview of a concrete example, explored in much more depth in the next section.

Definition 2.1: Let $\{e_n\}$ be an orthonormal basis of the Hilbert space \mathbb{H} . A sequence $\{h_n\} \in \mathbb{H}$ is a Riesz basis if there exists a bounded and boundedly invertible linear operator $\mathcal{R}: \mathbb{H} \mapsto \mathbb{H}$ such that $\mathcal{R}e_n = h_n$ for each n in some indexing set \mathbb{I} .

The symbol $\delta(\cdot)$ will denote the Dirac delta function and $\delta_{n,m}$ will be used to denote Kronecker’s delta so that, for example: $\langle f, \delta(\cdot - x) \rangle = f(x)$ for some $f \in \mathbb{H}$, and $\langle e_n, e_m \rangle = \delta_{n,m}$.

We will make use of the coefficient functionals associated with $\{e_n\}$ and $\{h_n\}$, namely $\mathcal{T}_e^*, \mathcal{T}_h^*: \mathbb{H} \mapsto \ell_2(\mathbb{I})$ defined by

$$\begin{aligned}\mathcal{T}_e^* \cdot &:= \{c_k(\cdot)\} := \{\langle \cdot, e_k \rangle\}_{k \in \mathbb{I}}, \\ \mathcal{T}_h^* \cdot &:= \{d_k(\cdot)\} := \{\langle \cdot, h_k \rangle\}_{k \in \mathbb{I}},\end{aligned}$$

and their adjoints $\mathcal{T}_e, \mathcal{T}_h: \ell_2(\mathbb{I}) \mapsto \mathbb{H}$, defined, for some $(a_k) \in \ell_2(\mathbb{I})$ by

$$\mathcal{T}_e a_k = \sum_{k \in \mathbb{I}} a_k e_k, \quad \mathcal{T}_h a_k = \sum_{k \in \mathbb{I}} a_k h_k.$$

We only consider complete orthogonal bases here. For then the Hilbert space is separable and we can write

$$h_n = \sum_{k \in \mathbb{I}} \langle h_n, e_k \rangle e_k = \sum_{k \in \mathbb{I}} c_k(h_n) e_k.$$

Throughout we shall focus on bases $\{h_n\}$ and $\{e_n\}$ which are ordered linearly with respect to scale in that:

$$h_n = h_1(n \cdot), \quad e_n = e_1(n \cdot).$$

We therefore have that $h_n(n^{-1} \cdot) = h_1 = \sum_{k \in \mathbb{I}} c_k(h_1) e_k$. Hence $h_n = \sum_{k \in \mathbb{I}} c_k(h_1) e_k(n \cdot)$, and using the linear ordering of $\{e_n\}$ gives

$$h_n = \sum_{k \in \mathbb{I}} c_k(h_1) e_n(k \cdot),$$

which reveals that the mapping $\mathcal{R}: e_n \mapsto h_n$ is equivalent to

$$\begin{aligned}h_n(x) &= \sum_{k \in \mathbb{I}} c_k(h_1) \langle \delta(\cdot - kx), \overline{e_n} \rangle \\ &= \sum_{k \in \mathbb{I}} c_k(h_1) (\overline{\mathcal{T}_e^* \delta(\cdot - kx)})(n). \quad (1)\end{aligned}$$

The main thrust of this work is to construct Riesz bases, subject to constraints, which approximate the behavior of an orthonormal basis in a local region of the transform domain. Again, assuming that $h_n = h_1(n \cdot)$, we have that the transform of a function $f \in \mathbb{H}$ with respect to the Riesz basis is:

$$\begin{aligned}(\mathcal{T}_h^* f)(n) &= \langle f, h_n \rangle = \sum_{k \in \mathbb{I}} c_k(h_1) \langle f, e_{kn} \rangle \\ &= \sum_{k \in \mathbb{I}} c_k(h_1) (\mathcal{T}_e^* f)(kn).\end{aligned}$$

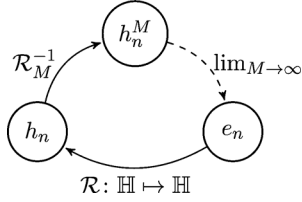


Fig. 1. Diagrammatic interpretation of the construction in (2). Terms are successively added such that the Riesz basis $\{h_n^M\}$ approaches the orthonormal basis $\{e_n\}$.

Here, we can see that, where $c_k(h_1)$ is non-zero, the ‘approximate’ transform $\mathcal{T}_h^* f$ at n is spread over the orthonormal transform domain at the k th multiples, or harmonics, of n ; or conversely, a non-zero coefficient in the orthogonal series will be present in the Riesz basis series, together with scaled, rational subharmonics. If one is to use the Riesz basis series to approximate the orthogonal series, then the error is

$$(\mathcal{T}_h^* f - \mathcal{T}_e^* f)(n) = \sum_{k \in \mathbb{I} \setminus \{1\}} c_k(h_1) (\mathcal{T}_e^* f)(kn).$$

We minimize this error locally by fusing what Candès [4] and others might regard a dictionary of functions or bases:

$$h_n: \mathbb{R} \times \Lambda \mapsto \mathbb{C}, \quad \forall n \in \mathbb{I},$$

over the set of construction parameters Λ . Let $M \in \mathbb{N}$. As illustrated in Fig. 1, of particular interest, here, are parametric superpositions of $\{h_{n,m}\}$ of the form $h_n^M: \mathbb{R} \times \Lambda^M \mapsto \mathbb{C}$, with $\Lambda^M := \prod_{m=1}^M \Lambda$, namely:

$$h_n^M := \sum_{m=-M}^M \gamma_m h_{n,m}, \quad h_{n,m} = h_n(\cdot, \lambda_m), \quad \lambda_m \in \Lambda, \quad (2)$$

such that $h_n^M \rightarrow e_n = \sum_{m \in \mathbb{Z}} \gamma_m h_{n,m}$ and where there exists a family of bounded and boundedly invertible linear operators $\mathcal{R}_M \rightarrow \mathcal{R}$ such that $\mathcal{R}_M h_n^M = h_n$, i.e., $h_n^M = \mathcal{R}_M^{-1} h_n \rightarrow \mathcal{R}^{-1} h_n = e_n$. It is in this sense that one approach could be to construct h_n^M , or equivalently, h_n and \mathcal{R}_M , such that the error

$$\|\mathcal{R}_M^{-1} h_n - e_n\|_{\mathbb{H}},$$

is minimized. Since \mathcal{T}_e^* is an isometry this is equivalent to minimizing $\|\mathcal{T}_e^* \mathcal{R}_M^{-1} h_n - \delta_{\cdot, n}\|_{\ell_p(\mathbb{I})}$. In fact, we are interested in a slight modification of this problem whereby the error is minimized over a local subset of the indexes using the norm $\|\cdot\|_{\ell_p(\mathbb{I}')}$, where $\mathbb{I}' \subseteq \mathbb{I}$. To this end, we define the objective, or cost, function as follows.

Definition 2.2: (Cost function) Let $\mathcal{T}_e, \mathcal{R}, h_n, \mathbb{I}'$ be defined as above. Then the cost, or objective, function associated with the basis $\{\mathcal{R}_M^{-1} h_n\}$ is defined by

$$C_p(\lambda; M, K) := \|\mathcal{T}_e^* \mathcal{R}_M^{-1} h_n - \delta_{\cdot, n}\|_{\ell_p(\mathbb{I}')} . \quad (3)$$

Note here that: the cardinality of the local indexes is $\#\mathbb{I}' = 2K$ to localize the approximation; the cardinality of the basis codomain is constrained to $2M$; and λ denotes a vector of parameters used in the construction of the basis $\{\mathcal{R}_M^{-1} h_n\}$.

Example 2.3: The top-left plot of Fig. 5, described in much more detail later on in Section VI, offers a very initial, and motivating example. The square-wave function h_1^1 , plotted in black

and with a codomain of $\{0, \pm 1\}$, is designed to approximate $\sin \pi t$. By optimizing the points at which the square wave transitions to and from the value zero it is possible to cancel out the third harmonic (i.e., to make the Fourier coefficients $c_k(h_1^1)$ equal to zero at $k = 3$; this can be seen in the resulting Fourier series, plotted in black, immediately below it. As such, this is an example where the cost function

$$\|c_k(h_1^1) - \delta_{k,1}\|_{\ell_p(\mathbb{I}')} ,$$

as measured in the local Fourier series domain, is zero for $k \in \mathbb{I}' = \{0, 1, 2, 3, 4\}$. In contrast, it is worth noting here that, although the grey colored plot in the same figures minimizes the two-norm distance between the square wave and the sinusoid over the larger interval $0 \leq j \leq 6$, it fails to offer the optimal solution over the more local interval $k \in \mathbb{I}'$. In this sense, the black-colored function may be considered ‘better’ than the grey-colored function. Likewise, the left-most plots of Figs. 6 and 7 and show that, as the cardinality of the approximating function is increased, so it is possible to systematically construct functions such that the sine wave is matched exactly over an increasingly large local interval of the Fourier domain.

III. PIECEWISE CONSTANT CONSTRUCTION

To explore more explicit realizations of the general problem posed above, we consider periodic, square integrable spaces on a finite interval. In this setting, and without loss of generality with respect to the period, we fix $\mathbb{H} = L_2(-\pi, \pi)$, with the orthonormal basis $e_n = e^{-in\cdot} / \sqrt{2\pi}$.

Definition 3.1: Let $f \in L_2(-\pi, \pi)$ be periodic with period 2π , then f can be expressed as the Fourier series:

$$f(x) \sim \frac{1}{\sqrt{2\pi}} \sum_{k \in \mathbb{Z}} c_k(f) e^{ikx},$$

with the Fourier coefficients $(c_k)_{k \in \mathbb{Z}} \in \ell_2(\mathbb{Z})$ defined by

$$c_k(f) := \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx.$$

Equivalently, $\mathcal{T}_e: c_k(f) \mapsto \sum_n c_k(f) e_n$ and $\mathcal{T}_e^*: f \mapsto \langle f, e_n \rangle$.

For $n \in \mathbb{Z}$, we construct the Riesz basis functions $\{h_n\}$ such that they share symmetry and periodicity properties with the set of exponentials $\{e^{in\cdot}\}$.

Definition 3.2: In what follows, define $h_n \in L_2(-\pi, \pi)$ such that the following hold almost everywhere.

- 1) h_1 is 2π periodic: $h_1 \equiv h_1(\cdot + 2\pi)$
- 2) h_n is Hermitian: $h_n \equiv \overline{h_n}(\cdot)$
- 3) $\{h_n\}$ is linearly ordered with respect to scale in that $h_n \equiv h_1(n\cdot)$

In the remainder of this section two different constructions will be given that satisfy Definition 3.2. Both take the general form of (2). The Möbius construction simply chooses $h_{n,m} = h_{nm}$. In fact, it is shown that instances of this construction have appeared in multiple works elsewhere. However, it will be shown that this is inefficient in terms of ‘dynamic range’, or the number of uniformly quantized levels required to represent the basis. As such this contrasts with, and motivates, the consideration of the ‘uniform construction’ which adds characteristic functions of uniform amplitudes and achieves a much more efficient dynamic range. First, the Möbius construction is described; then

the measurement of dynamic range is established. Finally, the uniform construction is presented.

A. Möbius Construction

It is instructive to briefly discuss a simple example taken from [17] and re-cast it into our framework. Originally, the construction was designed in the context of Paley-Wiener spaces and the application was that of multichannel sampling theory. The result is here placed into the context of our piecewise basis approximation setting. The centre-piece of the construction is the Möbius arithmetic function.

Definition 3.3: (Möbius arithmetic function) Let $\mu : \mathbb{N} \mapsto \{0, \pm 1\}$ denote the Möbius arithmetic function, given by

$$\mu(n) := \begin{cases} 1, & \text{if } n = 1 \\ (-1)^k, & \text{if } n \text{ is the product of } k \text{ distinct primes} \\ 0, & \text{otherwise} \end{cases}$$

Lemma 3.4: (Möbius function identity) Let μ denote the Möbius function. Then, for $k \in \mathbb{N}$,

$$\sum_{m|k} \mu(m) = \begin{cases} 1, & \text{for } k = 1 \\ 0, & \text{otherwise} \end{cases}$$

where $m | k$ denotes that m divides k .

Proof: See e.g., [2].

Proposition 3.5: (cf. [17]) Let the Riesz basis $\{h_n\}$ take the form of (2) and let it satisfy Definition 3.2 with $h_{n,m} := h_{nm}$, where the ratio $c_{km}(h_1)/c_k(h_1)$ is independent of k . Then $\exists \gamma_m$ such that $h_n^M \rightarrow e_n$ as $M \rightarrow \infty$.

Proof: We can find γ_m explicitly as follows. Note that $h_{nm}((nm)^{-1}) = h_1 = \sum_k c_k(h_1)e_k$, i.e., $h_{nm} = \sum_k c_k(h_1)e_{knm}$. Hence

$$h_n^M = \sum_{\substack{k,m \in \mathbb{Z} \\ |m| \leq M}} \gamma_m c_k(h_1) e_{knm}.$$

Now, setting

$$\gamma_m := \frac{c_{km}(h_1)}{c_k(h_1)} \mu(|m|),$$

gives

$$h_n^M = \sum_{k \in \mathbb{Z}} c_k(h_1) e_{kn} \sum_{\substack{m|k \\ |m| \leq M}} \mu(|m|).$$

Normalizing h_1 such that $c_1(h_1) = 1$ and an appeal to Lemma 3.4 completes the proof. ■

Remark 3.6: Define $\zeta_M : \mathbb{Z} \mapsto \mathbb{Z}$ by

$$\zeta_M(k) := \sum_{\substack{m|k \\ |m| \leq M}} \mu(|m|),$$

Then, we note that $\zeta_M(k) = 0, \forall k \in \mathbb{Z}_M := \{m \in \mathbb{Z} : |m| \leq M\} \setminus \{1\}$. Hence, similar to (1), we have

$$h_n^M = \sum_{k \in \mathbb{Z}} c_k(h_1) \zeta_M(k) \left(\overline{\mathcal{T}_e^* \delta(\cdot - kx)} \right)(n), \quad (4)$$

and, as $M \rightarrow \infty$, we have $\zeta_M(k) \rightarrow \delta_{k,1}$ and the support of the terms on the right-hand-side, namely $\mathbb{Z} \setminus \mathbb{Z}_M$ converges to $\{1\}$ and $h_n^M \rightarrow e_n$.

The following example has been considered in various guises and contexts, cf.: [10], [24]–[26], [29].

Example 3.7: Consider

$$h_n = \sqrt{\pi/32} (\text{sgn } \cos(n \cdot) + i \text{sgn } \sin(n \cdot)).$$

Then

$$c_m(h_1) = \begin{cases} m^{-1}, & m \in 4\mathbb{Z} + 1 \\ 0, & \text{otherwise} \end{cases}$$

and

$$h_n^M = \sum_{\substack{m \in 4\mathbb{Z} + 1 \\ |m| \leq M}} \frac{\mu(|m|)}{m} h_n(m \cdot).$$

For $M = 3 : h_n^3 = h_n + \frac{1}{3}h_n(-3 \cdot)$; for $M = 5 : h_n^5 = h_n + \frac{1}{3}h_n(-3 \cdot) - \frac{1}{5}h_n(5 \cdot)$; etc.

Example 3.7 gives a simple demonstration of a basis $\{h_n^3\}$, which is piecewise low-order (zero degree) polynomial and which shares the same coefficient functionals c_m of the (much smoother) Fourier basis over a localized set ($\{|m| = 1, 2, 3, 4\}$). Likewise $\{h_n^5\}$ does the same, but over the set $\{m \in \mathbb{Z} : |m| \leq 5\}$, etc. Together h_n^3, h_n^5, \dots can be seen as a family of Riesz basis constructions which converge to the orthogonal (Fourier) basis.

Although the Möbius construction may be the simplest way to approximate the Fourier basis with piecewise zero-degree polynomials, as explained in the next subsection, it is not necessarily the most ‘efficient’ in terms of dynamic range.

B. Dynamic Range

The dynamic range of a basis is the lowest common multiple of the denominators of the basis function codomain when normalized to the interval $[-1, 1]$ and expressed in lowest terms. The dynamic range determines how finely the normalized codomain has to be (uniformly) partitioned or quantized in order to describe the range of values that the basis takes. In the following, denote the real part of a complex number by \Re and the imaginary part by \Im . Denote the (codomain) restriction of a function to the positives by $|_+$ and restriction to the negatives by $|_-$, namely:

$$\begin{aligned} h(x)|_+ &:= \{(x, h(x)) : h(x) > 0\} \\ h(x)|_- &:= \{(x, h(x)) : h(x) < 0\}. \end{aligned}$$

Definition 3.8: Let $h_1 : \mathbb{R} \mapsto \mathbb{Q}(i)$ and let

$$\begin{aligned} \frac{\Re h_1(x)|_+}{\max_x |h_1(x)|} &= \frac{p^+(x)}{q^+(x)} \\ -\frac{\Re h_1(x)|_-}{\max_x |h_1(x)|} &= \frac{p^-(x)}{q^-(x)} \end{aligned}$$

where $p^+, p^-, q^+, q^- : \mathbb{R} \mapsto \mathbb{N}$, such that

$$\gcd(p^+(x), q^+(x)) = 1, \text{ and } \gcd(p^-(x), q^-(x)) = 1.$$

Then the dynamic range of h_1 is defined by

$$\text{dran } h_1 := \text{lcm}(\text{dran } \Re h_1, \text{dran } \Im h_1),$$

with

$$\text{dran} \Re h_1 := 2\text{lcm}(\text{dran} \Re h_1|_+, \text{dran} \Re h_1|_-),$$

where $\text{dran} \Im h_1$ is similarly defined, and

$$\begin{aligned} \text{dran} \Re h_1|_+ &:= \text{lcm}\{q^+(x)\} \\ \text{dran} \Re h_1|_- &:= \text{lcm}\{q^-(x)\}. \end{aligned}$$

Remark 3.9: Note that, for our purposes, since the negative part of the basis functions are simply shifted versions of the positive part, i.e., $e_n(\cdot + \pi) = -e_n$, our constructions are such that: $\text{dran} \Re h_1|_+ = \text{dran} \Re h_1|_-$. Furthermore, since the real and imaginary parts of the basis that we are attempting to approximate are simply shifted versions of each other, i.e., $(\Re e_n)(\cdot - \pi/2) = \Im e_n$, we need only be interested in constructions that satisfy $\text{dran} \Re h_1 = \text{dran} \Im h_1$. Hence, only the dynamic range of the positive and real part, say, need be computed to find the overall dynamic range, thus

$$\begin{aligned} \text{dran } h_1 &= \text{dran } \Re h_1 \\ &= 2\text{dran } \Re h_1|_+. \end{aligned}$$

Example 3.10: Consider a function $h: \mathbb{R} \mapsto \mathbb{Q}$ with

$$\frac{\Re h}{\max |h|} \in \{\pm 1, \pm 2/3, \pm 3/4\}$$

Then $q^+, q^- \in \{1, 3, 4\}$ which hence gives $\text{dran } h = 24$; that is, 24 uniformly spaced quantization levels are required to represent the function h up to a constant.

Example 3.11: For the constructions h_n^3 and h_n^5 in Example 3.7. We have $\text{dran } h_1^3 = 2\text{dran } \Re h_1^3|_+$. Now

$$\frac{\Re h_1^3|_+}{\max |h_1^3|} \in \{1, 1/2\},$$

which yields $q^+ \in \{1, 2\}$. Hence $\text{dran } \Re h_1^3|_+ = 2$ and $\text{dran } h_1^3 = 4$.

Likewise, $\text{dran } h_1^5 = 2\text{dran } \Re h_1^5|_+$ and

$$\frac{\Re h_1^5|_+}{\max |h_1^5|} \in \{1, 17/23, 7/23\},$$

whence we have $q^+ \in \{1, 23\}$, $\text{dran } \Re h_1^5|_+ = 23$, and $\text{dran } h_1^5 = 46$. On the other hand, the cardinality of the (non-zero values of the) codomains of h_n^3 , h_n^5 are, respectively 4 and 8. That it should take 46 quantization levels to represent only 8 different values motivates the construction in Section III.C where the dynamic range is equal to the codomain cardinality (cf. power-of-two filters [14]). In other words, the dynamic range is optimal; constraints on the dynamic range are therefore equivalent to constraints on the cardinality of the codomain.

C. Uniform Construction

We modify the form in (2) and, for each m , put $\gamma_m = 1$. We drop the assumption that $h_{n,m} = h_n(m \cdot)$ and find an alternative to the design described by Proposition 3.5. Unlike the

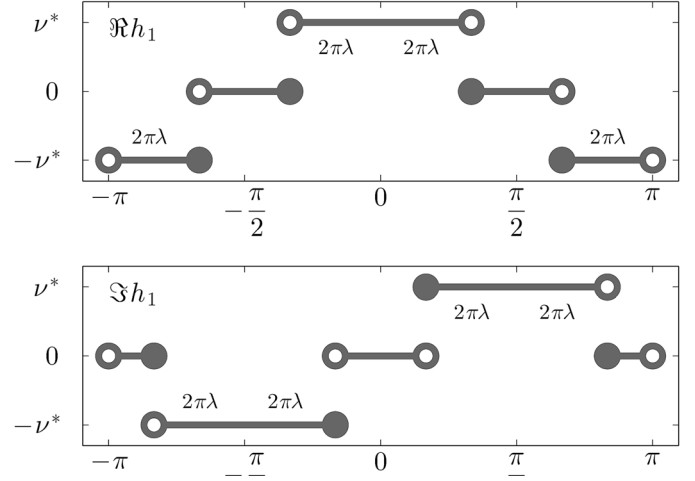


Fig. 2. An example of the real (top) and imaginary (bottom) parts of h_1 from Definition 3.12. For convenience, we have put $\nu^* := \nu\sqrt{\pi/32}$. The $2\pi\lambda$ annotations indicate the widths of the constituent line elements.

original treatment of this construction ([18]) we here place it into the Riesz basis selection framework. These results are then exploited in the next section to establish a connection with a constrained binary optimization problem.

Definition 3.12: Let $h_{n,m} = h_n(\cdot, \lambda_m)$ be piecewise constant, complex functions in $L_2(-\pi, \pi)$ which are periodic and scale-ordered (as in Definition 3.2), such that, for $x \in (-\pi, \pi)$

$$\begin{aligned} (\Re h_1)(x) &= \begin{cases} 1, & x \in (-2\pi\lambda_m, 2\pi\lambda_m) \\ -1, & x \in (-\pi, \pi(2\lambda_m - 1)] \cup [\pi(1 - 2\lambda_m), \pi) \\ 0, & \text{otherwise} \end{cases} \\ (\Im h_1)(x) &= \begin{cases} 1, & x \in [\pi(1/2 - 2\lambda_m), \pi(1/2 + 2\lambda_m)) \\ -1, & x \in (\pi(-1/2 - 2\lambda_m), \pi(-1/2 + 2\lambda_m)] \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

with $\lambda_m \in (0, 1/4)$. Then the uniform step basis (or uniform construction) is defined by

$$h_n^M := \sum_{m=1}^M h_n(\cdot, \lambda_m), \quad c_k(h_1^M) = 1.$$

An example of the construction elements h_1 are plotted in Fig. 2.

Remark 3.13: Note that, akin to the exponential basis e_n , the real and imaginary parts of the construction h_1 are designed to be even and odd, respectively.

Theorem 3.14: Let $\{h_n^M\}$ be the uniform construction and let $\mathcal{T}_e^* = \{\langle \cdot, e_k \rangle\}_k$. Then, for some $\{c_k \zeta_M(k)\} \in \ell_2(\mathbb{Z})$,

$$h_n^M(x) = \sum_{k \in 4\mathbb{Z}+1} c_k \zeta_M(k) \left(\overline{\mathcal{T}_e^* \delta(\cdot - kx)} \right) (n).$$

Proof: Define the double-indexed set of intervals

$$\Delta_n^j = \begin{cases} (\pi(-1/2 - 2\lambda_m), \pi(-1/2 + 2\lambda_m)], & j = -1 \\ (-2\pi\lambda_m, 2\pi\lambda_m), & j = 0 \\ [\pi(1/2 - 2\lambda_m), \pi(1/2 + 2\lambda_m)), & j = 1 \\ (-\pi, \pi(2\lambda_m - 1)] \cup [\pi(1 - 2\lambda_m), \pi), & j = 2 \end{cases}$$

Denote the indicator, or characteristic, function by χ is $\chi: (-\pi, \pi) \mapsto \{0, 1\}$, namely

$$\chi_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}.$$

For then the construction elements can be written

$$h_1(\cdot, \lambda_m) := \nu \sqrt{\frac{\pi}{32}} \left((\chi_{\Lambda_m^0} - \chi_{\Lambda_m^2}) + i \left(\chi_{\Lambda_m^1} - \chi_{\Lambda_m^{-1}} \right) \right), \quad (5)$$

where the normalization factor $\nu \in \mathbb{R}$, is introduced to ensure $c_1(h_1^M) = 1$. We have

$$\begin{aligned} c_k(h_1(\cdot, \lambda_m)) &= \frac{\nu}{8} \left(\int_{\Lambda_m^0} - \int_{\Lambda_m^2} + i \left(\int_{\Lambda_m^1} - \int_{\Lambda_m^{-1}} \right) \right) e^{-ikx} dx \\ &= \begin{cases} \nu k^{-1} \sin 2k\pi \lambda_m, & k \in 4\mathbb{Z} + 1 \\ 0, & \text{otherwise} \end{cases}. \end{aligned} \quad (6)$$

Now, note that $h_{n,m}(n^{-1}\cdot) = h_n(n^{-1}\cdot, \lambda_m) = h_1(\cdot, \lambda_m) = \sum_k c_k(h_{1,m})e_k$ and hence that

$$h_{n,m} = \sum_k c_k(h_{1,m})e_{kn},$$

and we have

$$h_n^M = \sum_{m=1}^M \sum_{k \in 4\mathbb{Z}+1} c_k(h_{1,m})e_{kn}.$$

Note that $c_k(h_1(\cdot, 1/4)) = k^{-1}$. Hence, defining $\zeta_M(k) := \sum_{m=1}^M \sin 2k\pi \lambda_m$, gives

$$\begin{aligned} h_n^M(x) &= \sum_{k \in 4\mathbb{Z}+1} c_k(h_1(\cdot, 1/4)) \zeta_M(k) e_{kn}(x) \\ &= \sum_{k \in 4\mathbb{Z}+1} c_k(h_1(\cdot, 1/4)) \zeta_M(k) \left(\overline{\mathcal{T}_e^* \delta(\cdot - kx)} \right)(n) \end{aligned} \quad (7)$$

which is of the specified form with $c_k = c_k(h_1(\cdot, 1/4))$. ■

Remark 3.15: We note that (7) is similar to (4) and that, once again, we require $\zeta_M(k) = 0$, $\forall k \in \mathbb{Z}_K$, for some $K \in \mathbb{N}$; i.e., cf. (6), we want to choose the lengths $\lambda = (\lambda_m) \in (0, 1/4)^M$ such that

$$\sum_{m=1}^M \sin 2k\pi \lambda_m = 0, \quad (8)$$

for all $|k| \leq K$ and $k \neq 1$.

In [18], it was shown that a solution to $\zeta_M(k) = 0$ can be found analytically, in terms of the Sylvester-type Hadamard matrix.

Theorem 3.16: (cf. [18]) Let M^* be the $(M_0 + 1)$ th prime and let $M = 2^{M_0}$. Then a solution λ to

$$\sum_{m=1}^M \sin 2k\pi \lambda_m = 0, \quad \forall k = 2, 3, \dots, M^*,$$

$$\lambda = \mathbf{H}_{M_0} \ell, \quad (9)$$

with $\mathbf{H}_0 = 1$, and

$$\mathbf{H}_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

and, in general, $\mathbf{H}_{M_0} = \bigotimes_{m=1}^{M_0} \mathbf{H}_1$, where \bigotimes denotes the tensor, or Kronecker, product, and where $\ell_1 = (k'_1 + 1)/6$ and

$$\ell_j = \begin{cases} \frac{2k'_j + 1}{4k_j}, & j \text{ power of two} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

and where $\{k_j\}$ are the set of odd ascending primes and $\{k'_j\} \in \mathbb{Z}$.

Proof: See [18].

Remark 3.17: Note that the Hadamard-Sylvester solution (9) not only ensures that all terms in the sum of (7), or equivalently in (8), are cancelled for $2 \leq k \leq M^*$ but that they are also cancelled for all multiples of $2 \leq k \leq M^*$.

Example 3.18:

- 1) Setting $M_0 = 0, k'_1 = 0$, gives $\ell = \lambda_1 = 1/6$ and (7) and (8) are zero for $k = 3$.
- 2) Setting $M_0 = 1, k'_1, k'_2 = 0$, gives $\ell = [1/6, 1/20]^T \Rightarrow \lambda = [13, 7]^T/60$ and (7) and (8) are zero for $k = 3, 5$.
- 3) Setting $M_0 = 2, k'_j = 0 \forall j$, gives $\ell = [1/6, 1/20, 0, 1/28]^T \Rightarrow \lambda = [53, 32, 38, 17]^T/210$ and (7) and (8) are zero for $k = 3, 5, 7$. Note that $53/210 > 1/4$ but we wanted $\lambda_m \in (0, 1/4)$. However, cf. Fig. 2 and note that, from Definition 3.2, the $h_1(\cdot, \lambda_m)$ are 2π -periodic; hence, adding a construction element with width $53/210$ is equivalent to adding a width of $52/210$ (the overlapping segments cancel).
- 4) Setting $M_0 = 2, k'_1, k'_2, k'_3 = 0, k'_4 = 1$, gives $\ell = [1/6, 1/20, 0, 3/28]^T \Rightarrow \lambda = [68, 2, 23, 47]^T/210$ and (7) and (8) are zero for $k = 3, 5, 7$. Note that $68/210 > 1/4$ but we wanted $\lambda_m \in (0, 1/4)$. However, again noting 2 and Definition 3.2, adding a construction element with width $68/210$ is equivalent to adding a width of $37/210$ (the overlapping segments cancel).

IV. MINIMUM CARDINALITY FRAMEWORK

The previous section described a zero-degree basis construction with small dynamic range. In this section Corollary 4.1 below confirms that this is also a solution to an instance of the optimization problem stated in Definition 2.2.

A. Relationship Between the Minimum Cardinality Cost Function and the Hadamard-Sylvester Solutions

Recall that the goal is to approximate elements of the orthonormal basis, over a local subset of their ‘analysis domain’ (e.g., over a local subset of Fourier coefficients) with superpositions of piecewise polynomial Riesz basis functions, subject to cardinality constraints on the codomain.

Now recall the objective function given by Definition 2.2. For a basis, $\{\psi_n\}$ say, that satisfies Definition 3.2, we want to minimize the cost function subject to cardinality constraints on the scaled basis codomain:

$$\min_{\#\psi=2M+1} C_p(\lambda; M, K).$$

By construction, the uniform step basis has $\#h_n^M = 2M + 1$ and is normalized (by the constant ν) such that $(\mathcal{T}_e^* h_1^M)(1) = 1$. Hence, for the uniform construction, we want to minimize

$$C_p(\lambda; M, K) = \|\mathcal{T}_e^* h_1^M\|_{\ell_p(\mathbb{Z}_K)}, \quad h_1^M = h_1^M(\cdot, \lambda), \quad (11)$$

with respect to $\lambda \in (0, 1/4]^M$ and where we recall the notation used earlier: $\mathbb{Z}_K := \{k \in \mathbb{Z} : |k| \leq K\} \setminus \{1\}$. In other words, we construct the optimal h_1^M by minimizing the objective function and then simply use property (3) of Definition 3.2 to set $h_n^M = h_1^M(n \cdot), \forall n$. Ostensibly, this can be seen as an integer valued, cardinality-constrained, filter bank version of the sparsity constrained, finite impulse response construction problem discussed by Baran *et al.* [3]. The previous section showed an analytical solution to a very similar cost minimization problem and we can now state this connection as follows.

Corollary 4.1: Define the cost function, cf. Definition 2.2

$$C_p(\lambda; M, K) := \|\mathcal{T}_e^* \mathcal{R}_M^{-1} h_n - \delta_{\cdot, n}\|_{\ell_p(\mathbb{I})}.$$

Let $M = 2^{M_0}$ and let M^* denote the $(M_0 + 1)$ th largest prime. Let \mathbf{H}_M be the M th order Hadamard-Sylvester matrix and let ℓ be defined as in (10). Then the Hadamard-Sylvester solution $\mathbf{H}_{M_0} \ell$ also solves

$$\min_{\lambda} C_p(\lambda; M, M^*).$$

Proof:

$$\begin{aligned} C_p(\lambda; M, M^*) &= \|\langle h_1^M, e \cdot \rangle\|_{\ell_p(\mathbb{Z}_{M^*})} \\ &= \left\| \left\langle \sum_{n \in 4\mathbb{Z}+1} c_n \zeta_M(n) e_n, e \cdot \right\rangle \right\|_{\ell_p(\mathbb{Z}_{M^*})} \\ &= \left\| \sum_{n \in 4\mathbb{Z}+1} c_n \zeta_M(n) \langle e_n, e \cdot \rangle \right\|_{\ell_p(\mathbb{Z}_{M^*})} \\ &= \|c \cdot \zeta_M(\cdot)\|_{\ell_p(\mathbb{Z}_{M^*})}, \end{aligned}$$

where the first equality follows from the definition of the cost function, cf. (11), the second equality results from Theorem 3.14, and the fourth equality is a consequence of the orthonormality of $\{e_n\}$. We then note, from Theorem 3.16, that the Hadamard-Sylvester solution gives $\zeta_M(k) = 0 \forall k \in \mathbb{Z}_{M^*} \Rightarrow C_p(\lambda; M, M^*) = 0$. ■

B. Relationship Between the Minimum Cardinality Cost Function and Constrained Binary Optimization

Since the problem described in (11) involves piecewise constant functions, we can frame it as a constrained binary optimization problem, as follows.

Theorem 4.2: Let $h_n^M = h_n^M(\cdot, \lambda)$ be the uniform construction with associated cost function $C_p(\lambda; M, K)$. Then there exists a matrix $\mathbf{A} \in \mathbb{C}^{(2K+1) \times N}$ and binary vector $\mathbf{x}_\lambda \in \{0, 1\}^N$ such that

$$C_p(\lambda; M, K)_{\ell_p(\mathbb{Z}_K)} = \|\mathbf{A} \mathbf{x}_\lambda\|.$$

Proof: We note firstly that, by virtue of the Fourier series derivative property $(\mathcal{T}_e \mathcal{D}f)(k) = ik(\mathcal{T}_e f)(k)$, the objective function becomes

$$\begin{aligned} C_p(\lambda; M, K) &= \|(\cdot)^{-1} \mathcal{T}_e^* \mathcal{D} h_1^M\|_{\ell_p(\mathbb{Z}_K)} \\ &= \left(\sum_{k \in \mathbb{Z}_K} \left| \frac{1}{k} \left(\mathcal{T}_e^* \frac{d}{dx} h_1^M(x, \lambda) \right)(k) \right|^p \right)^{1/p}, \end{aligned} \quad (12)$$

where the derivatives should be interpreted in the distributional sense. For small enough $\Delta = (N-1)^{-1}/4$, $N \in \mathbb{N}_*$, we can write (the piecewise constant function) h_1^M in discretized form as the vector

$$\mathbf{h}_\lambda^M \propto (h_1^M(\pi(8n\Delta - 1), \lambda))_{n=0}^{N-1} \in \mathbb{C}^N,$$

i.e., the construction lengths are discretized as $\lambda \in \Lambda_\Delta^M := \{(n\Delta)_{n=1}^N\}^M$. For then, the operator \mathcal{T}_e^* becomes an appropriately scaled, and truncated, discrete Fourier transform matrix \mathbf{W}_K with elements proportional to $e^{-i\pi n \Delta k/N}$, for $n = 0, \dots, N-1$ and $k = -K, \dots, K$. Hence, the objective function is

$$C_p(\lambda; M, K) = \|\mathbf{W}_K \mathbf{h}_\lambda^M\|_{\ell_p(\mathbb{Z}_K)},$$

with constraint $\mathbf{w}^T \mathbf{h}_\lambda^M = 1$ where $\mathbf{w} \propto (e^{-i\pi n \Delta/N})_{n=0}^{N-1}$. Here, the elements in the matrix \mathbf{W}_K are scaled (all by the same constant factor) such that the vector \mathbf{h}_λ^M takes integer values. Hence, (12) becomes

$$C_p(\lambda; M, K) = \|\Omega \mathbf{W}_K \nabla \mathbf{h}_\lambda^M\|_{\ell_p(\mathbb{Z}_K)},$$

where ∇ is the (circular) discrete difference operator with $\nabla a_k := a_k - a_{k-1}$ for $k = 2, \dots, K$ and $\nabla a_1 := a_K - a_1$; and where

$$\Omega := \text{diag}(|k^{-1}|)_{k=-K}^K.$$

We note here that $\nabla \mathbf{h}_\lambda^M$ is a vector with elements in $\{0, \pm 1\}$. However, by virtue of Definition 2.3, since h_n^M is Hermitian (the real part is even and the imaginary part is odd) we can multiply the necessary real and imaginary parts of the coefficients in the Fourier matrix, say $\mathbf{W}_K \mapsto \widetilde{\mathbf{W}}_K$, and define the (circular) absolute difference operator $\nabla_+ \cdot := |\nabla \cdot|$. We then have

$$C_p(\lambda; M, K) = \|\Omega \widetilde{\mathbf{W}}_K \nabla_+ \mathbf{h}_\lambda^M\|_{\ell_p(\mathbb{Z}_K)}.$$

Now, $\nabla_+ \mathbf{h}_\lambda^M$ is a vector with elements in $\{0, 1\}$, and the minimization is a constrained optimization problem, subject to the constraint that $\nabla_+ \mathbf{h}_\lambda^M$ is binary and that $\mathbf{w}^T \mathbf{h}_\lambda^M = 1$. ■

C. Relationship Between the Constrained Binary Optimization and the Hadamard-Sylvester Solutions

We therefore arrive at the interesting result that the Hadamard-Sylvester solutions intersect with a subset of binary optimization solutions.

Corollary 4.3: Let $M = 2^{M_0}$ and let M^* denote the $(M_0 + 1)$ th largest prime. The Hadamard-Sylvester solution $\lambda = \mathbf{H}_{M_0} \ell$ (cf. Theorem 3.16) solves the binary optimization problem

$$\min_{\mathbf{x} \in \{0,1\}^N} \|\mathbf{A}\mathbf{x}\|_{\ell_p(\mathbb{Z}_{M^*})},$$

subject to $\mathbf{w}^T \mathbf{h}_\lambda^M = 1$, with $\mathbf{A} = \Omega \widetilde{\mathbf{W}}_{M^*}$ and $\mathbf{x} = \nabla_+ \mathbf{h}_\lambda^M$.

Proof: Result follows immediately from Corollary 4.1 and Theorem 4.2. ■

Remark 4.4: The parameter K in the matrix $\widetilde{\mathbf{W}}_K$ controls the size of the subset of frequency components around $k = 0$ (i.e., ‘base-bandwidth’) over which we compare the Riesz basis with the orthonormal basis. That these components decay as $O(|k|^{-1})$ serves as motivation to minimize the distance between the bases over finite K . In this sense, it is tempting to also parameterize the matrix Ω by introducing a decay exponent α to control the weight or importance of the components near to zero. We then arrive at the following generalization of the objective function, namely

$$C_p(\lambda; M, K, \alpha) = \left\| \Omega^\alpha \widetilde{\mathbf{W}}_K \nabla_+ \mathbf{h}_\lambda^M \right\|_{\ell_p(\mathbb{Z}_K)}. \quad (13)$$

V. EXTENSIONS TO HIGHER ORDER POLYNOMIALS

The piecewise constant construction can be extended to piecewise polynomials in at least two ways, namely antiderivatives and convolutions. These are treated in the next two subsections below.

A. Antiderivative Family

Definition 5.1: Let $\{h_1^{M,1}\} := \{h_n^M\}$ be the uniform construction given above. Let $\{h_n^{M,q}\}$ satisfy Definition 3.2, with

$$h_1^{M,q}(x) \propto C_q + \int_{-\pi}^x h_1^{M,q-1}(\xi) d\xi, \quad x \in [-\pi, \pi]$$

such that $c_1(h_1^{M,q}) = 1$ and where the constant $C_q \in \mathbb{R}$, defined by

$$C_q := \int_{-\pi}^{\pi} \int_{-\pi}^x h_1^{M,q-1}(\xi) d\xi dx,$$

ensures that $\int_{-\pi}^{\pi} h_n^{M,q}(x) dx = 0$.

Remark 5.2: Note that $h_1^{M,q}$, over q , is the family of $(q-1)$ th antiderivatives of h_1^M :

$$\mathcal{D}^{q-1} h_n^{M,q} = h_n^M.$$

Furthermore, we have that

$$\mathcal{T}_e^* \mathcal{D}^q h_n^{M,q} = (ik)^{q-1} \mathcal{T}_e^* \mathcal{D} h_n^M.$$

We can now consider the problem of constructing the piecewise polynomials $h_n^{M,q}$ to locally approximate $\{e_n\}$, i.e., we generalize the cost function from (12) to the form

$$C_p^q(\lambda; M, K) = \left\| \mathcal{T}_e^* h_1^{M,q} \right\|_{\ell_p(\mathbb{Z}_K)} = \left\| (\cdot)^{-q} \mathcal{T}_e^* \mathcal{D} h_1^M \right\|_{\ell_p(\mathbb{Z}_K)}.$$

But, this is equivalent to raising the weighting matrix Ω in (13) to the power of q and we have that

$$C_p^q(\lambda; M, K, \alpha) = C_p(\lambda; M, K, q\alpha).$$

Hence, optimizing the construction of q order piecewise polynomials is equivalent to performing a constrained binary optimization of the piecewise constant construction with a weight matrix decay of q . The examples given in Fig. 3 show that, as expected, these functions approximate the orthonormal basis (of complex exponentials $\{e_n\}$) with increasing accuracy as either the number M of construction elements is increased and/or as the order $q-1$ of the polynomial is increased.

B. Convolution Family

Definition 5.3: Let $\{h_n^M\}$ be the uniform construction. Let $\{\eta_n^{M,q}\}$ satisfy Definition 3.2, with $\eta_1^M := \eta_1^{M,1} := h_1^M$. Then, define

$$\eta_1^{M,q} := \eta_1^{M,q-1} * \eta_1^M := \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \eta_1^{M,q-1}(x) \eta_1^M(\cdot - x) dx.$$

Remark 5.4: Note that $\mathcal{T}_e^* \eta_n^{M,q} = (\mathcal{T}_e^* \eta_n^M)^q$.

Hence, the alternative way to construct piecewise polynomials to locally approximate $\{e_n\}$ is to minimize

$$\begin{aligned} \Gamma_p^q(\lambda; M, K) &:= \left\| \mathcal{T}_e^* \eta_1^{M,q} \right\|_{\ell_p(\mathbb{Z}_K)} \\ &= \left\| \mathcal{T}_e^* h_1^M \right\|_{\ell_{pq}(\mathbb{Z}_K)}^q \\ &= |C_{pq}(\lambda; M, K)|^q. \end{aligned}$$

With weights, this is

$$\begin{aligned} \Gamma_p^q(\lambda; M, K, \alpha) &:= \left\| \Omega^\alpha \widetilde{\mathbf{W}}_K \nabla_+ \eta^M \right\|_{\ell_p(\mathbb{Z}_K)} \\ &= \left\| \Omega^\alpha \widetilde{\mathbf{W}}_K \nabla_+ \mathbf{h}_\lambda^M \right\|_{\ell_{pq}(\mathbb{Z}_K)}^q \\ &= |C_{pq}(\lambda; M, K, \alpha)|^q. \end{aligned}$$

Hence, optimizing this alternative construction of q order piecewise polynomials is equivalent to performing a constrained binary optimization of the piecewise constant construction with respect to an ℓ_{pq} , rather than an ℓ_p , norm.

Similar to Fig. 3, the examples given in Fig. 4 show that these extensions approximate the orthonormal basis $\{e_n\}$ with increasing accuracy as either the number M of construction elements is increased and/or as the order $q-1$ of the polynomial is increased.

VI. EXPERIMENTS

Corollary 4.3 gives the result that the Hadamard-Sylvester (analytical) solution coincides with the solution to the Fourier basis version of the binary optimization problem $C_p(\lambda; M, K)$ for $K = M^*$. In addition to the proofs given in the previous

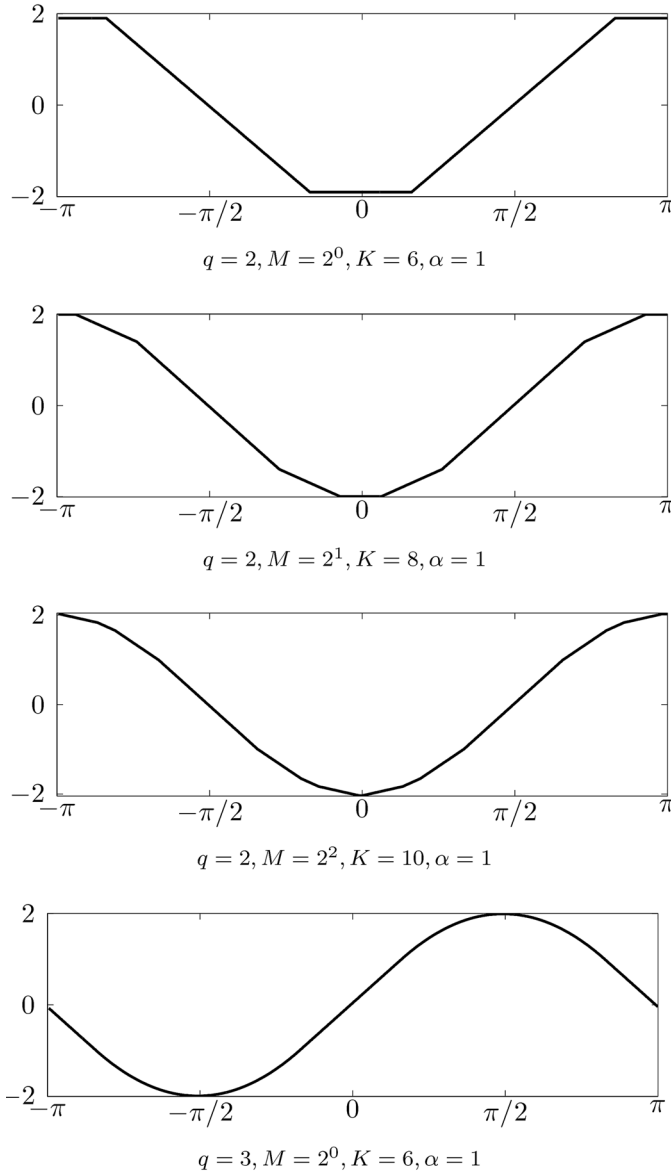


Fig. 3. Antiderivative extension to order $q-1=1$ and $q-1=2$ polynomials. As either the number of ‘levels’ M , or the order q is increased, the functions become more sinusoidal.

section, this is now corroborated experimentally. Of added interest, here, is to investigate the numerical solutions for the case $K > M^*$. That is, whether a numerical solution can cancel out more harmonics than the Hadamard-Sylvester solution or whether the analytical solution is as good as it gets in this regard.

Tables I, II, III show the numerically optimized values of the length parameter λ for $M = 2^0, 2^1, 2^2$ respectively. These are computed, via an exhaustive search, with $\Delta = 1/420$, for several weight decays $\alpha = 0, 1, 2, 4$, and for several bandwidths $K = 4, 6, \dots, 22$. The Δ is chosen in this way to ensure that the search intervals are discretized sufficiently finely that they contain the exact Hadamard-Sylvester solutions given by Theorem (3.16). Figs. 5, 6, 7 illustrate the ‘optimal’ basis functions (in the sense of (13)) found by the numerical minimization for a selection of different decays and bandwidths.

In particular, the tables show that the Hadamard solution given by (9) is found by the ℓ_1 minimizer for: $M = 2^0, K = 4$;

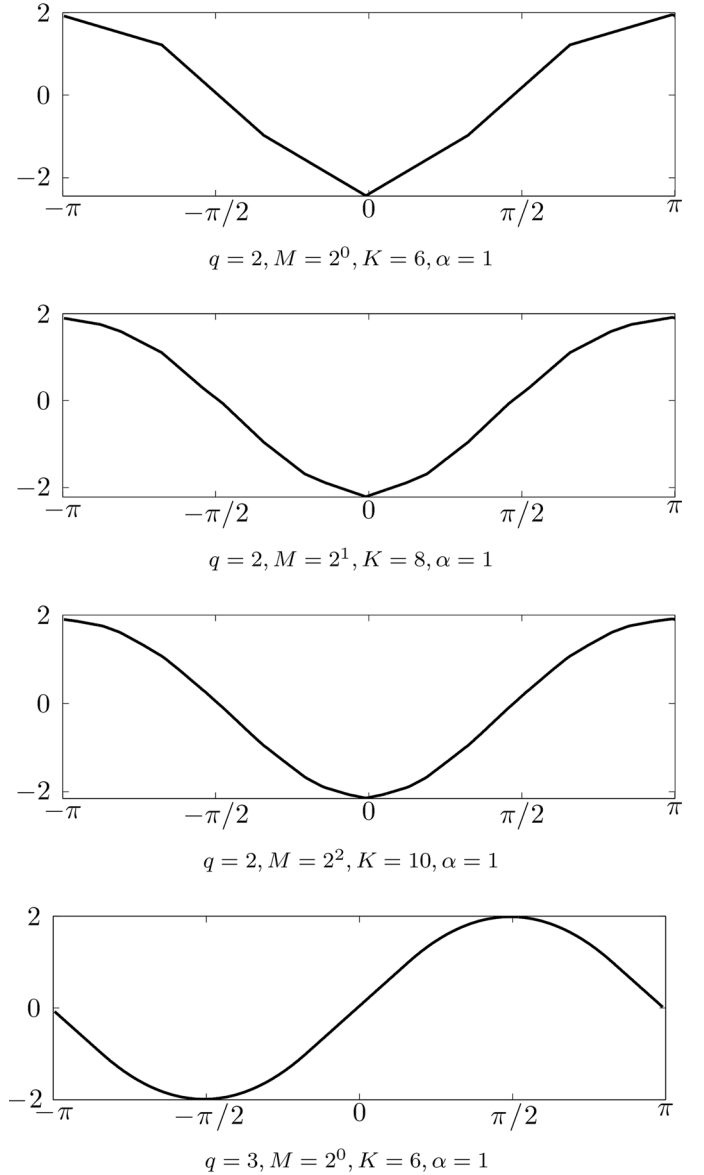


Fig. 4. Convolution extension to order $q-1=1$ and $q-1=2$ polynomials. As either the number of ‘levels’ M , or the order q is increased, the functions become more sinusoidal.

$M = 2^1, K = 6$; $M = 2^2, K = 8$. This confirms our previous analytic findings that we can zero out: the 3rd harmonic of h_n^1 ; the 3rd and 5th of h_n^2 ; and the 3rd, 5th, and 7th harmonics of h_n^4 . The numerical results also show that we cannot do better than the analytic solution; that is, for $M = 1$, we cannot zero out both the 3rd and 5th harmonic; or, for $M = 2$ we cannot zero out the 3rd, 5th, and 7th harmonic.

We can also see that analytic solution prevails in some cases where the bandwidth is increased (the Hadamard solution appears in more than one row and column in the tables). Furthermore, this holds for ℓ_1 solution more than it does for the ℓ_2 solution. This reflects two things: (1) the harmonics of the construction elements h_n decay as $1/k$ so that the lower harmonics contain most of the ‘error’; (2) ℓ_2 minimization tends to lead to solutions which spread the error over several series coefficients whereas the ℓ_1 solution, is better at inducing ‘sparsity’ in the series coefficients. On the other hand, increasing the

TABLE I

OPTIMAL LENGTH PARAMETERS (λ) FOR VARIOUS WEIGHT DECAYS α AND BANDWIDTHS K FOR THE $M = 2^0$ CASE. THE TOP FOUR ROWS ARE THE ℓ_1 SOLUTIONS; THE BOTTOM FOUR ARE THE ℓ_2 SOLUTIONS. (THE ENTRIES BELOW ARE ALL FRACTIONS OF 420; NOTE, FROM EXAMPLE 3.18, PART 1, THE HADAMARD SOLUTION: $\lambda = 1/6 = 70/420$)

$\alpha \backslash K$	4	6	8	10	12	14	16	18	20	22
0	70	84	84	90	93	95	96	97	98	98
1	70	70	84	70	70	70	70	70	70	70
2	70	70	70	70	70	70	70	70	70	70
4	70	70	70	70	70	70	70	70	70	70
0	70	81	87	90	93	94	96	97	98	98
1	70	77	80	74	76	78	79	76	77	78
2	70	73	72	72	72	72	72	72	72	72
4	70	70	70	70	70	70	70	70	70	70

TABLE II

OPTIMAL LENGTH PARAMETERS (λ) FOR VARIOUS WEIGHT DECAYS α AND BANDWIDTHS K FOR THE $M = 2^1$ CASE. THE TOP FOUR ROWS ARE THE ℓ_1 SOLUTIONS; THE BOTTOM FOUR ARE THE ℓ_2 SOLUTIONS. (THE ENTRIES BELOW ARE ALL FRACTIONS OF 420; NOTE, FROM EXAMPLE 3.18, PART 2, THE HADAMARD SOLUTION: $\lambda = [13, 7]/60 = [91, 49]/420$)

$\alpha \backslash K$	4	6	8	10	12	14	16	18	20	22
0	(71,69)	(91,49)	(81,39)	(85,55)	(91,61)	(92,69)	(94,74)	(96,77)	(96,80)	(97,83)
1	(71,69)	(91,49)	(91,49)	(91,49)	(85,55)	(91,49)	(91,49)	(89,51)	(87,53)	(87,53)
2	(78,8)	(91,49)	(91,49)	(91,49)	(91,49)	(91,49)	(91,49)	(91,49)	(91,49)	(91,49)
4	(78,8)	(91,49)	(91,49)	(91,49)	(91,49)	(91,49)	(91,49)	(91,49)	(91,49)	(91,49)
0	(91,21)	(91,49)	(83,40)	(87,53)	(90,62)	(92,68)	(94,73)	(95,76)	(96,79)	(97,82)
1	(91,21)	(91,49)	(88,52)	(88,52)	(90,57)	(91,61)	(91,53)	(91,53)	(90,55)	(89,54)
2	(91,21)	(91,49)	(89,51)	(89,51)	(89,52)	(89,51)	(89,51)	(89,51)	(89,51)	(89,51)
4	(91,21)	(91,49)	(90,50)	(90,50)	(90,50)	(90,50)	(90,50)	(90,50)	(90,50)	(90,50)

weight decay can also help the minimizer find the Hadamard solution for lower order ($M = 1, 2$) constructions; this is because the decay penalizes the lower harmonics more than the higher harmonics.

In Fig. 5, a selection of ‘optimal’ uniform constructions are plotted (on the interval $[-\pi, \pi]$), alongside (for illustrative purposes) the Fourier series coefficients of their derivative. (Note, that the Fourier series of the non-differentiated function is simply $1/k$ multiplied by the series of the derivative.) The first column shows the optimal ℓ_1 solution for $K = 6$, i.e., over the Fourier coefficients $|k| \leq 6$; this particular case coincides with the Hadamard solution given by Example 3.18, part 1, (which gives a minimum cost function of zero) whereas the ℓ_2 norm solution does not. The ℓ_1 solution induces sparsity in the $k = 3$ coefficient at the cost of a relatively larger error in the $k = 5$ coefficient. In this sense, the ℓ_1 solution obtains a better ‘local’ approximation to $\{e_n\}$.

In the second column, we see that, by widening the ‘bandwidth’ to $K = 8$, neither norms produce the same minimum as the Hadamard solution. It is clear that a zero minimum does not exist in this case. Furthermore, from the Fourier series plot, we can see that the ℓ_2 solution spreads the error between $k = 3, 5$, and 7 ; whereas, the ℓ_1 solution zeros out the Fourier coefficient at $k = 5$ and spreads the error between $k = 3$ and 7 .

The sparsity inducing effect of increasing the weight decay can be seen in the third column. In this case, the Hadamard solution is obtained (exactly for ℓ_1 and approximately for ℓ_2) over coefficients $|k| \leq 22$.

TABLE III

OPTIMAL LENGTH PARAMETERS (λ) FOR VARIOUS WEIGHT DECAYS α AND BANDWIDTHS K FOR THE $M = 2^2$ CASE. THE TOP FOUR ROWS ARE THE ℓ_1 SOLUTIONS; THE BOTTOM FOUR ARE THE ℓ_2 SOLUTIONS. (THE ENTRIES BELOW ARE ALL FRACTIONS OF 420; NOTE, FROM EXAMPLE 3.18, PARTS 3 AND 4, THE HADAMARD SOLUTIONS: $\lambda = [104, 76, 64, 34]/420$ AND $\lambda = [94, 74, 46, 4]/420$)

$\alpha \backslash K$	4	6	8	10	12	14	16	18	20	22
0	(78,71 8,1)	(104,76 64,34)	(104,76 64,34)	(104,76 64,34)	(96,71 43,2)	(96,81 38,61)	(100,81 58,39)	(98,79 61,42)	(95,78 61,45)	(95,79 61,45)
1	(78,71 8,1)	(97,71 69,27)	(104,76 64,34)	(104,76 64,34)	(96,72 44,2)	(96,80 62,35)	(100,81 59,40)	(98,79 61,42)	(98,79 61,42)	(98,79 61,42)
2	(78,71 8,1)	(80,74 10,4)	(104,76 64,34)	(104,76 64,34)	(94,74 46,4)	(96,80 62,35)	(96,80 62,35)	(96,80 62,35)	(96,80 62,35)	(96,80 62,35)
4	(78,71 8,1)	(80,74 10,4)	(94,74 46,4)	(94,74 46,4)	(94,74 46,4)	(94,74 46,4)	(94,74 46,4)	(94,74 46,4)	(94,74 46,4)	(94,74 46,4)
0	(91,78 21,8)	(80,74 10,4)	(94,74 46,4)	(94,74 46,4)	(95,70 41,3)	(96,80 37,60)	(97,78 56,34)	(98,81 61,42)	(95,76 57,39)	(96,79 61,45)
1	(91,78 21,8)	(80,74 10,4)	(94,74 46,4)	(94,74 46,4)	(95,71 2,42)	(96,81 61,37)	(98,37 60,80)	(97,80 61,39)	(96,78 60,38)	(97,80 62,41)
2	(91,78 21,8)	(80,74 10,4)	(94,74 46,4)	(94,74 46,4)	(96,81 61,36)	(96,81 61,36)	(97,80 61,36)	(98,36 97,61)	(97,80 62,36)	(97,80 62,36)
4	(91,78 21,8)	(80,74 10,4)	(94,74 46,4)	(94,74 46,4)	(96,80 62,35)	(96,80 62,35)	(96,80 62,35)	(96,80 62,35)	(96,80 62,35)	(96,80 62,35)

Figs. 6 and 7 also show the optimal solutions (in the sense that they remove the 3rd and 5th or 3rd, 5th, and 7th harmonics respectively) found by both norms in the first column and other solutions in the second and third column. Interestingly, the first column of Fig. 7 shows that the ℓ_1 and ℓ_2 norms find two different Hadamard solutions; ℓ_1 finds Example 3.18, part 3 and ℓ_2 finds Example 3.18, part 4. Indeed, the ℓ_2 solution might be deemed marginally better; the basis function contains more zeros and the first non-zero Fourier series coefficient at $k = 11$ is smaller than that of the ℓ_1 solution. (Regrettably, we cannot offer a satisfactory explanation for this.)

VII. CONCLUSION AND FURTHER WORK

The work presented above shows how minimal cardinality Riesz basis approximations to orthogonal bases can be constructed by inducing sparsity in the distributional derivatives. We have shown that this is equivalent to a constrained binary optimization problem and, for the case of the Fourier series basis, we have found that the solutions intersect those found with Sylvester-type Hadamard matrices [18]. However, following a remark made in Section IV, the first column of Fig. 7 poses the question of whether it is possible to prove which Hadamard solution corresponds to the ℓ_1 solution and which one corresponds to the ℓ_2 solution? Furthermore, since both solutions are optimal, how do we choose between them? Other possible further work is briefly outlined, as follows.

The discussed framework affords the possibility of generalizing the sparse filter construction of Baran *et al.* [3] to the filter-bank, integer valued case. For example, if we require a low-cardinality approximation to a Riesz basis $\{\psi_n\}$ over some indexing subset $\mathbb{I}' \subset \mathbb{I}$, then the cost function in Definition 2.2 becomes: $\|(\mathcal{T}_e^* \mathcal{R}_M^{-1} h_n)(\cdot) - (\mathcal{T}_e^* \psi_n)(\cdot)\|_{\ell_p(\mathbb{I})}$. The construction elements h_n would, of course, have to be modified from (5), depending on what kind of Hilbert space the $\{\psi_n\}$ are defined in. A particular challenge would be whether an analytical solution exists (cf. the Hadamard solution for the uniform construction) and how it would be found. As an aside, reinterpreting

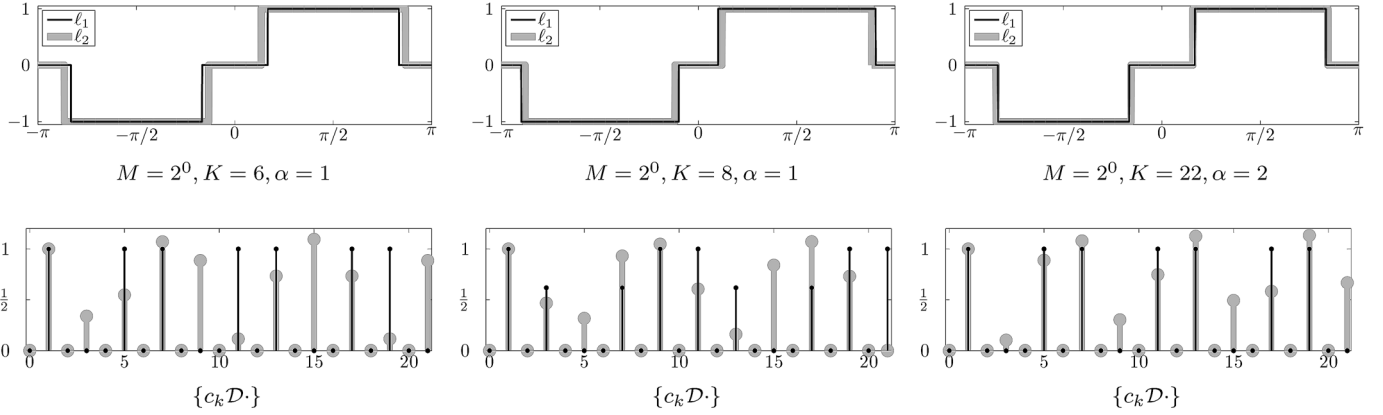


Fig. 5. Uniform constructions for $M = 1$ and a selection of ‘bandwidths’, K , and weight decays α . The (imaginary part of the) ‘time-domain’ function is shown in the first row and the Fourier series of the derivative in the second row.

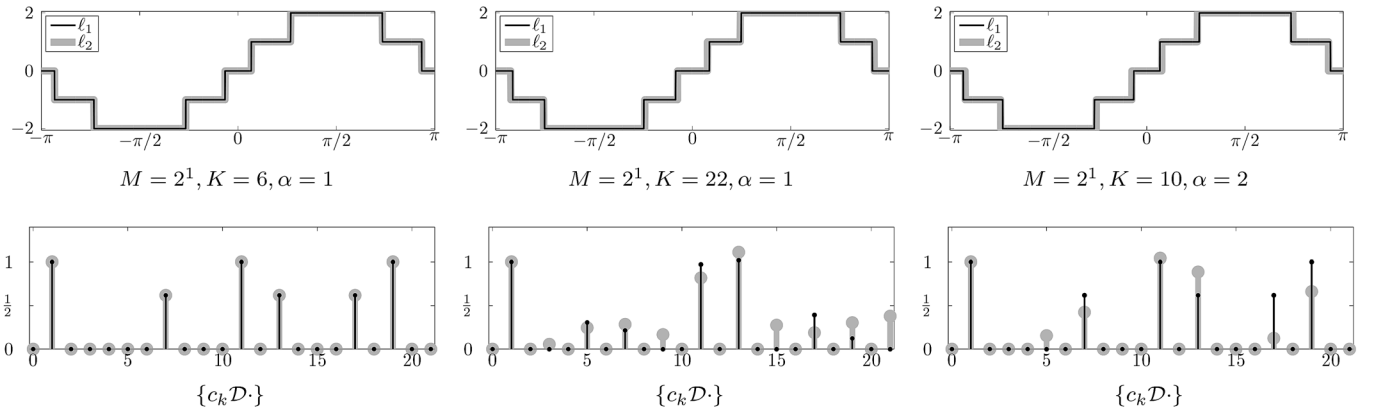


Fig. 6. Uniform constructions for $M = 2$ and a selection of ‘bandwidths’, K , and weight decays α . The (imaginary part of the) ‘time-domain’ function is shown in the first row and the Fourier series of the derivative in the second row.

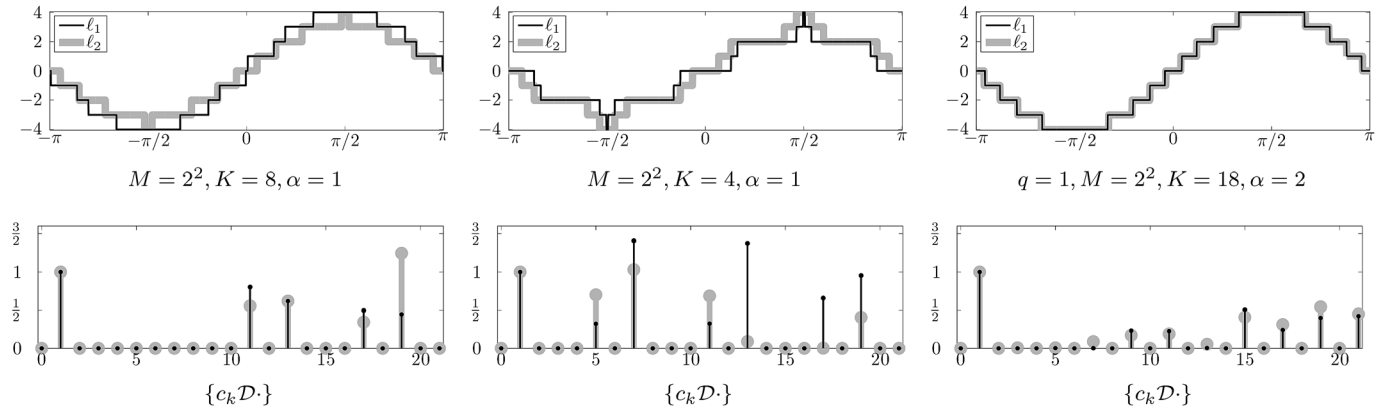


Fig. 7. Uniform constructions for $M = 4$ and a selection of ‘bandwidths’, K , and weight decays α . The (imaginary part of the) ‘time-domain’ function is shown in the first row and the Fourier series of the derivative in the second row.

the Hilbert space as a Paley-Wiener space gives rise to a simple instance of the compressive sensing type problem, namely that of minimizing a cost over the discrete sampled time domain, subject to sparsity constraints in the frequency domain.

Motivated by the approximate signal processing philosophy of Nawab and Dorken [16], an iterative refinement algorithm was proposed in [18]. This decomposed the uniform, piecewise constant construction such that the resulting transform could be performed iteratively. However, this scheme was purely deterministic. Instead, since the error will depend on the complexity

of the signal, it would be of interest to explore a probabilistic approach, cf. Winograd and Nawab [28], to predict the signal dependent algorithmic complexity in an adaptive manner.

The Mobius construction plays a central part of both the wavelet filter enhancement scheme of [19] and the multi-channel, multisampling rate theorem of [17]. It is tempting, therefore, to explore a development of these ideas using, instead, the uniform construction.

Iserles and Nørsett [11] introduce a modification to the Fourier transform (a shift in the imaginary part of the basis

functions) which induces a stronger decay in the Fourier series. Very much in contrast, the work presented here focuses on the behavior of our modified transform over the coefficients c_k for $k \in \mathbb{Z}_K$ rather than the decay as $k \rightarrow \infty$. Notwithstanding, it should be possible to combine both of these ideas and explore the idea of introducing a shift in the imaginary part of the uniform, piecewise constant basis. As Iserles and Nørsett show, however, that justifying this seemingly simple shift with sufficient rigor requires much care and is also, therefore, left as further work.

ACKNOWLEDGMENT

The author is most grateful for the reviewers' detailed comments and suggestions. In addition, the author would like to extend special thanks to Kelly Watson–Nelson for her comments and kind support throughout the preparation of this paper.

REFERENCES

- [1] D. Anastasia and Y. Andreopoulos, "Software designs of image processing tasks with incremental refinement of computation," *IEEE Trans. Imag. Process.*, vol. 19, no. 8, pp. 2099–2114, 2010.
- [2] T. M. Apostol, *Introduction to Analytic Number Theory*. New York, NY, USA: Springer, 1976.
- [3] T. Baran, D. Wei, and A. V. Oppenheim, "Linear programming algorithms for sparse filter design," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1605–1617, 2010.
- [4] E. J. Candès, "New ties between computational harmonic analysis and approximation theory. approximation," *Innovations Appl. Math.*, pp. 87–153, 2002.
- [5] E. J. Candès, Y. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *J. Appl. Comput. Harmon. Anal.*, vol. 31, pp. 59–73, 2010.
- [6] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Comput.*, vol. 20, pp. 33–61, 1998.
- [7] M. Gaster and J. B. Roberts, "On the spectral analysis of randomly sampled records by a direct transform," *Proc. Roy. Soc. A*, vol. 354, pp. 27–58, 1977.
- [8] T. I. Haweel and A. Alhasan, "A simplified square wave transform for signal processing," in *An International Conference on Mathematical Analysis and Signal Processing*, ser. Contemporary Mathematical Series. Providence, RI, USA: Amer. Math. Soc., 1994, pp. 265–271.
- [9] H. Hedenmalm, P. Lindqvist, and K. Seip, "A Hilbert space of Dirichlet series and systems of dilated functions in $L^2(0, 1)$," *Duke Math. J.*, vol. 86, pp. 1–37, 1997.
- [10] R. D. Hughes and M. L. Heron, "Approximate Fourier transform using square waves," *Proc. Inst. Electr. Eng. —A*, vol. 136, pp. 223–228, 1989.
- [11] A. Iserles and S. P. Nørsett, "From high oscillation to rapid approximation I: Modified fourier expansions," *IMA J. Numer. Anal.*, vol. 28, pp. 862–887, 2008.
- [12] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *J. Appl. Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 234–253, 2001.
- [13] M. P. Lamoureux, "The Poorman's transform: Approximating the Fourier transform without multiplication," *IEEE Trans. Signal Process.*, vol. 41, no. 3, pp. 1413–1415, 1993.
- [14] D. Li, Y. C. Lim, Y. Lian, and J. Song, "A polynomial-time algorithm for designing FIR filters with power-of-two coefficients," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1935–1941, 2002.
- [15] V. Lohweg, C. Diederichs, and D. Müller, "Algorithms for hardware-based pattern recognition," *EURASIP J. Appl. Signal Process.*, vol. 12, pp. 1912–1920, 2004.
- [16] S. Nawab and E. Dorken, "Efficient STFT approximation using a quantization and differencing method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1993, vol. 3, pp. 587–590.
- [17] J. D. B. Nelson, "A multi-channel, multi-sampling rate theorem," *Sampl. Theory Signal Image Process.: An Int. J.*, vol. 2, pp. 83–96, 2003.
- [18] J. D. B. Nelson, "On the coefficient quantization of the Fourier basis," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1838–1846, 2003.
- [19] J. D. B. Nelson, "A wavelet filter enhancement scheme with a fast integral B-wavelet transform and pyramidal multi-B-wavelet algorithm," *J. Appl. Comput. Harmon. Anal.*, vol. 18, pp. 234–251, 2005.
- [20] J. Pender and D. Covey, "New square wave transform for digital signal processing," *IEEE Trans. Signal Process.*, vol. 40, no. 8, pp. 2095–2097, 1992.
- [21] J. E. T. Penny, M. I. Friswell, and D. J. Inman, "Approximate frequency analysis in structural dynamics," *Mechan. Syst. Signal Process.*, vol. 27, pp. 370–378, 2012.
- [22] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, 2005.
- [23] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *J. Appl. Comput. Harmon. Anal.*, vol. 3, no. 2, pp. 186–200, 1996.
- [24] D. W. Tufts and G. Sadasiv, "The arithmetic Fourier transform," *IEEE ASSP Mag.*, vol. 5, pp. 13–17, 1988.
- [25] W. J. Walker, "Sampling theory and the Arithmetic Fourier Transform," in *Sampling Theory in Fourier and Signal Analysis: Advanced Topics*, J. R. Higgins and R. L. Stens, Eds. New York, NY, USA: Oxford Univ. Press, 1999, ch. 2, pp. 38–55.
- [26] Y. Wei, "Dirichlet multiplication and easily-generated function analysis," *Comput. Math. Appl.*, vol. 39, pp. 173–199, 2000.
- [27] Y. Wei, "Frequency analysis based on easily generated functions," *J. Appl. Comput. Harmon. Anal.*, vol. 8, pp. 280–285, 2000.
- [28] J. Winograd and S. H. Nawab, "Probabilistic complexity analysis for a class of approximate DFT algorithm," *J. VLSI Signal Process.*, vol. 14, pp. 193–205, 1996.
- [29] A. Wintner, *An Arithmetical Approach to Ordinary Fourier Series*. Baltimore, MD, USA: Waverly Press, 1945.



James D. B. Nelson joined the Department of Statistical Science at University College London as a Lecturer (Assistant Professor) in 2010 and became Senior Lecturer (Associate Professor) in 2013. After a Ph.D. in applied harmonic analysis from the Mathematics Department at Anglia Polytechnic University (1998–2001), he held post-doc positions in: the Applied Mathematics and Computing Group at the University of Cranfield (2001–2004); the Information: Signals, Images, and Systems Research Group at the University of Southampton (2004–2006); and the Signal Processing and Communications Laboratory at the University of Cambridge (2006–2010).

Dr. Nelson's interests span aspects of harmonic analysis, signal processing, computational statistics, and machine learning.